# Mixing Patterns Between Age Groups Using Social Networks

S. Del Valle<sup>1\*</sup>, J. M. Hyman <sup>2</sup>, H.W. Hethcote<sup>3</sup>, and S. G. Eubank<sup>1†</sup>

<sup>1</sup> CCS-5 Simulation Science (MS M997), Los Alamos National Laboratory,

Los Alamos, NM 87545, USA

<sup>2</sup> Mathematical Modeling Analysis Group (MS B284), Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>3</sup>Department of Applied Mathematical & Computational Sciences, University of Iowa, 14 McLean Hall, Iowa City, IA 52242, USA

#### Abstract

We describe a new method for estimating adequate-contact matrices that describe the mixing and the probability of disease transmission between age groups. Adequate-contact matrices can be used to estimate age-dependent forces of infection in age-structured, compartmental models for the study of the transmission dynamics of infectious diseases. The interactions among individuals determine the course of an epidemic through a population. To capture these elements of reality, we use the social network for the synthetic city of Portland, Oregon to study and identify mixing patterns. The mixing within the population consists of two groups, children and adults. Children interact with other children close to their own age, while adults interact with a wider range of age groups and their durations of a typical contact are shorter than a typical contact between children. Our

<sup>\*</sup>Author for correspondence (sdelvall@lanl.gov).

<sup>&</sup>lt;sup>†</sup>Present address: Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, USA

results show that children are more likely to have adequate contacts than adults. Understanding mixing in the population is a key factor in using mathematical modeling to better understand disease spread and the best means of containing an outbreak.

### 1 Introduction

A major determinant in understanding the spread of diseases is our lack of data on the mixing patterns in the population. It is important to appropriately account for the formation of contacts to accurately understand disease spread and develop control measures. Correctly accounting for the mixing patterns in a population may be crucial to accurately predict the path of a disease and thus where the outbreak could be intercepted most effectively.

Epidemiology models can estimate the likelihood of a disease outbreak based on the reproductive number. The reproductive number is the average number of secondary cases produced by a "typical" infectious individual during its infectious period (van den Driessche & Watmough 2002). The rate at which infectious individuals spread the disease depends on the number of adequate contacts between infecteds and susceptibles. Thus, if we determine the mixing patterns in the population, we can obtain better estimates of the reproductive number. This result can help modelers predict the severity of an outbreak and the best means of containing it.

Because of the recognition that heterogeneous contact patterns govern sexually transmitted diseases (STDs), several techniques have been developed to incorporate heterogeneous mixing in mathematical models for STDs. Numerous models have studied the effects of different mixing

functions or mixing matrices in the form of compartmental models (Anderson et al. 1990; Blythe & Castillo-Chavez 1989; Hyman & Stanley 1988, 1989; Hyman & Li 1996, 1997; Knolle 2004) and network models (Newman 2002; Zaric 2002). Some of the techniques developed to incorporate non-random mixing into epidemic models include restricted mixing (Jacquez et al. 1988), proportionate mixing (Hethcote & Van Ark, 1987; Nold 1980), preferred mixing (Hethcote & Yorke 1984), selective mixing (Koopman et al. 1989), and non-proportionate mixing (Anderson & May 1991). These techniques involve defining an  $n \times n$  matrix, the elements of which represent adequate contacts between individuals in age group i and age group i, where adequate contacts are those that would result in the transfer of infection. However, these matrices require knowledge of the forces of infection, the mixing structure, and the steady states of the endemic disease. The forces of infection are usually estimated using serological data, but these data are often not available for many diseases.

Survey studies of mixing patterns can be useful tools in understanding disease spread. Edmunds et al. (1997) studied a sample of 65 individuals and estimated contact patterns that could lead to the spread of airborne infections. They concluded that older adults mix with themselves and all other age groups at the same rate, and that younger adults do not. They also found that people have a different mixing pattern during the weekend than on weekdays. However, some of the limitations of this study are the sample size, the lack of quantification of duration of contact, and the fact that all the participants were adults, even though a great number of diseases are transmitted by children. A review article by Wallinga et al. (1999) discussed the use of networks in developing contact patterns and the spread of

airborne infectious diseases. They noted that more studies are needed to better understand contact patterns to predict disease spread.

We use a social network generated by the synthetic population of Portland, Oregon, consisting of more than 1.6 million individuals to determine mixing patterns between age groups. We analyze the simulated movement of these individuals and determine the likelihood of infection. The data used in this study such as number of contacts and duration of contact exhibits aspects of behavior thought to be important in determining patterns of infection. Furthermore, based on this data we estimate an adequate-contact matrix which is used in mathematical models to determine age-dependent forces of infection.

Our results show that the population is divided mainly into two blocks of mixing, school children and adults. Adults mix with a wider range of age groups than school children and teenagers. Moreover, the adequate-contact matrix demonstrates that the probability of disease transmission is higher for children than for adults, possibly due to the long duration of contacts at schools. Therefore, one implication of our results is that if there is an epidemic outbreak, closing schools could be a potentially efficient control strategy.

### 2 SOCIAL NETWORKS

Recently, considerable interest has been concentrated in determining the effect that social networks can have on understanding the spread of disease. Social networks try to explain the linkages among social entities and the implications of these linkages. The concept of a network emphasizes the fact that each individual has ties to other individuals, each of whom in turn

is tied to a few, some or many others. Using a network model approach in which individuals are the nodes and their contacts are the links, we show how data on interactions of a synthetic population can be translated into matrices and the influence of their structure on disease spread. Individual-based simulation is also being used increasingly to help epidemiological investigations. In agent-based simulation, the individual entities in the model are represented directly and possess an internal state and set of behaviors or rules which determine how the agent's state is updated from one time-step to the next. Here, we use both of these tools to examine mixing patterns in disease spread.

#### 3 EPIDEMIOLOGICAL SIMULATION SYSTEM

## (EpiSims)

EpiSims is an individual-based system for simulating the spread of disease in a large urban population (Barret et al. 2005; Chowell et al. 2003; Eubank et al. 2004). The original EpiSims model was based on the city of Portland, Oregon, in which the simulated movement of more than 1.6 million individuals was constructed. Information such as demographics and daily activities for the population were assigned based on the 2000 census data, activity surveys, their location and their distance to other locations. The city of Portland was mapped into 180,000 physical locations such as households, schools and shopping centers. This information was used to calculate the number of people at each location at each time in order to estimate contacts among individuals.

The simulation determines the contacts among individuals, including identities of those in contact, the location and duration of the contact, and the nature of the activity. Because for some locations such as large buildings, thousands of individuals may go to the same location, EpiSims creates an ad-hoc model for mixing within a location. In this "sub-location" model, people are assigned to rooms based on activity type. Assignments overlap, so a "worker" may have contact with a student. EpiSims integrates all this information into a computer model in order to provide estimates of physical contact patterns for a large human population.

#### 3.1 Population

EpiSims uses a synthetic population that resembles the real population of Portland, Oregon in the course of carrying out their daily activities over one randomly chosen day. The population consists of 1,615,860 individuals of ages ranging from 0 to 90 years. Figure 1 shows a histogram of the age distribution of the population. Portland is somewhat unusual because of the disproportionately large population of young adults, resulting in a double-hump distribution.

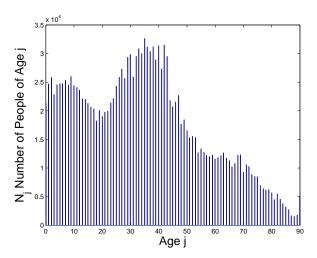


Figure 1: Age distribution of the synthetic population for the city of Portland. The population consists of 1,615,860 individuals of ages ranging from 0 to 90 years. The population is described by a double-hump distribution with mean of 34.37 and median of 33.

#### 3.2 Daily Number of Contacts

The social contact network of Portland is implemented as a directed network, where the nodes represent locations and the directed edges represent movement of individuals between locations. Each individual in the population is assigned an ID and second-by-second movement of each person is recorded. The EpiSims social network output includes the total number of contacts for each individual, the contacts' IDs, duration of each contact, and the location where each contact took place.

The population of Portland had a total of 27,613,194 contacts, ranging between 0 to 365 contacts per person per day. The average number of contacts per person in each age group is shown in figure 2. The average number of people contacted per person can give us an

estimate of how many secondary cases can potentially acquire infection from one index case. The average number of contacts generated by the synthetic population of Portland is consistent with previous studies (Edmunds *et al.* 1997).

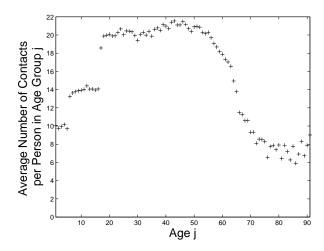


Figure 2: Average number of contacts per person in age group j.

We estimated the total number of contacts between age groups,  $C_{ij}$ , by using each persons' ID from the Episims output and matching the IDs with their demographic data and expressing it in a matrix form. The total number of contacts between age groups is illustrated in figure 3. Notice that this matrix is symmetric, so that  $C_{ij} = C_{ji}$  for all i and j. Figure 3 shows that the mixing in the population consists of two blocks of contacts: young individuals (< 20 years) and adults (> 20 years). We observe a weak coupling between children and adults, possibly due to child-parent contacts. Notice that adults mix more at random and with a wider range of age groups than young individuals.

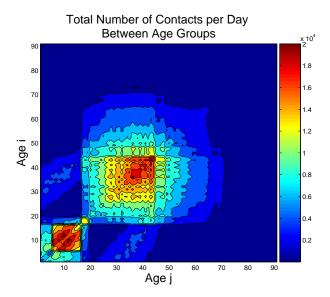


Figure 3: Total number of contacts  $C_{ij}$  between age groups. The mixing among the population is divided into two blocks of contacts, young individuals (< 20 years) and adults (> 20 years). We observe a weak coupling between children and adults, possibly due to child-parent contacts. Notice that adults mix more at random and with a wider range of age groups than young individuals.

Furthermore, we obtained the average number of daily contacts of a person in age group i with people in age group j, by dividing the total number of contacts  $C_{ij}$ , by the total population size  $N_i$  in age group i. The resulting  $n \times n$  matrix is defined as  $\gamma_{ij}$ .

#### 3.3 Duration of Contact

Closeness and duration of contact are important determinants of disease transmission. Surprisingly, none of the models used in the literature to study disease spread have quantified duration of contact. We used EpiSims output to determined the daily average duration of contact per

pair.

First, we estimate the total duration of all contacts between age groups,  $D_{ij}$  (figure 4). Notice that this matrix is also symmetric, so that  $D_{ij} = D_{ji}$  for all i and j. The duration of contact between age groups is consistent with the total number of contacts found in figure 3. Second, we estimate the average duration of contact,  $T_{ij}$ , by dividing the total duration of all contacts by the total number of contacts. That is,  $T_{ij} = (1/24)(D_{ij}/C_{ij})$ , where  $D_{ij}$  is the total duration of all contacts and  $C_{ij}$  is the total number of contacts between age groups. We divided  $T_{ij}$  by 24 to get the average duration of contacts in fractions of a day instead of in hours.

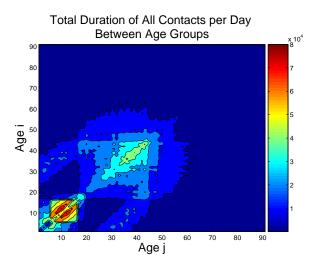


Figure 4: The total duration  $D_{ij}$  of all contacts between age groups.

### 4 ESTIMATING THE ADEQUATE-CONTACT

### **MATRIX**

Here we demonstrate how the EpiSims results can be used to estimate an adequate-contact matrix. The force of infection  $\lambda_i$  is the relative rate at which susceptibles of age i acquire infection. Homogeneous mixing means that contacts of a person are randomly distributed among all others in the population. One immediate implication of the assumption of homogeneous mixing is that the force of infection is same for all ages. However, for heterogeneous mixing, the forces of infection reflect the age-related differences in the degree of mixing and contact, within and among age groups.

Empirical evidence of age-related differences in  $\lambda_i$  have been documented for childhood infections by Anderson & May (1982, 1991) and Grenfell & Anderson (1985). They estimated forces of infection as a function of age using serological data or records of case notifications. These studies suggest that the age-related differences in the force of infection are important factors in modeling infectious diseases. The most direct evidence against the homogeneous mixing assumption comes from studies showing that for human diseases the force of infection  $\lambda_i$ , tends to increase with age up to about 5-15 years, and then to decrease in later years (Anderson & May 1991).

The standard method used in mathematical models to take account of age-dependent mixing patterns of the population is to use a WAIFW (Who Acquires Infection From Whom) matrix (Anderson & May 1991). The WAIFW matrix describes how individuals mix with other age

groups. The elements of the WAIFW matrix,  $\beta_{ij}$ , represent the rate at which an infective of age j will infect a susceptible of age i. However, this technique requires knowledge of the forces of infection, the mixing structure, and the steady states of the endemic disease. Nevertheless, the forces of infection are estimated using serological data, but this data are often not available for many diseases (Anderson & May 1991). Furthermore, the pre-judgment of the mixing structure may be unrealistic, so that the data leads to mixing matrices with negative entries. Therefore, there is a great need to develop new methods to estimate age-dependent forces of infection.

We use the social network for the city of Portland to estimate age-dependent forces of infection and, consequently, adequate-contact matrices. We assume that the population is demographically divided into different age groups that can progress through various infection stages (Hyman  $et\ al.\ 1999$ ). For this model, we consider 90 age groups and m infection stages.

We define the force of infection  $\lambda_i$  as the rate of disease transmission from infected people in all age groups to susceptibles in age group i. That is,  $\lambda_i$  is the sum of the rate of disease transmission from all infection stages in all age groups for age groups  $1 \leq j \leq 91$  and infection stages  $1 \leq k \leq m$ , to the susceptible group,  $S_i$ . This means that a susceptible person in age group i can get infected by a person in any infection stage in any age group. Thus,

$$\lambda_i = \sum_{j=1}^{91} \sum_{k=1}^m \lambda_{ijk}(t). \tag{1}$$

Where,  $\lambda_{ijk}$  is the rate of disease transmission from the infected people  $I_{jk}$  in infection stage k of age group j to the susceptibles in age group i. We calculate  $\lambda_{ijk}$  in (1) as the product of the number of contacts per unit time that each individual in age group i has with age group j;

the probability of disease transmission per contact between an infected in stage k of age group j and a susceptible in age group i; and the fraction of those contacts that are infected. That is,

$$\lambda_{ijk} = \begin{pmatrix} \text{Number of} \\ \text{Contacts per} \\ \text{Unit Time} \end{pmatrix} \begin{pmatrix} \text{Probability of} \\ \text{Disease Transmission} \\ \text{per Unit Time} \end{pmatrix} \begin{pmatrix} \text{Fraction of} \\ \text{Contacts that} \\ \text{are Infected} \end{pmatrix}.$$

In terms of the EpiSims data, we can define the force of infection  $\lambda_{ijk}$  as the product of the average number of contacts,  $\gamma_{ij}$ ; the probability of disease transmission, which is the product of the susceptibility  $(\alpha_i)$  of a susceptible in age group i, the infectivity  $(\xi_{jk})$  of an infective in stage k of age group j, and the probability of transmission  $P_{ij}$  based on the average duration of contacts between age groups i and j; and the fraction of contacts that are infected. That is,

$$\lambda_{ijk} = \left(\gamma_{ij}(t)\right) \left(\alpha_i \xi_{jk} P_{ij}\right) \left(\frac{I_{jk}(t)}{N_j(t)}\right), \tag{2}$$

where  $I_{jk}$  is the number of people in infection stage k of age group j and  $N_j$  is the size of age group j.

Let  $\sigma$  be the mean number of transmission events per hour of contact between fully infectious and fully susceptible people. For events that occur randomly in time, the number of occurrences in a period of time of length t obeys a Poisson probability law with parameter  $\sigma t$ . Thus, the probability of no occurrences in time interval t is  $e^{-\sigma t}$  and the probability of at least one occurrence is  $1 - e^{-\sigma t}$ . Using the mean duration  $T_{ij}$  of contacts between a person in age group i with people in age group j, we assume that the probability of transmission in this time interval

 $T_{ij}$  is given by

$$P_{ij} = 1 - e^{-\sigma T_{ij}}. (3)$$

Using Equation (3) with  $\sigma = 8$  and the average durations of contact per pair,  $T_{ij}$ , we obtain the probability of transmission for all age groups (figure 5). Figure 5 gives us the probability of transmission,  $P_{ij}$ , based on the average duration of contacts between age groups. Because of the larger average durations of contact among people of the same age, the probabilities of transmission are high along the diagonal for all age groups.

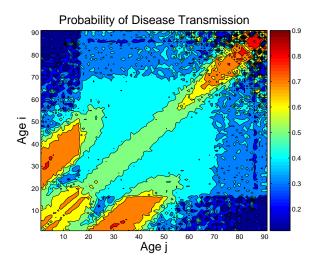


Figure 5: Probability of transmission  $P_{ij}$  between age groups. We observe a high probability of transmission along the diagonal for all age groups.

Lastly, we define  $\beta_{ij}$  as the average number of adequate contacts (a contact that is sufficient for transmission) between a susceptible in age i with people in age j, which is the product of the average number of contacts, the susceptibility, the infectivity, and the probability of disease transmission, that is,  $\beta_{ij} = \gamma_{ij} * \alpha_i * \xi_{jk} * P_{ij}$ . Using  $\gamma_{ij}$ ,  $\alpha_i = 1$ ,  $\xi_{jk} = 1$ , and  $P_{ij}$ , we

estimate an adequate-contact matrix  $\beta_{ij}$  (figure 6), for the given social network. For simplicity, we assume in this example that all age groups are equally susceptible ( $\alpha_i = 1$ ), and that all infected individuals are equally infectious ( $\xi_{jk}$ ), regardless of the infection stage or age group.

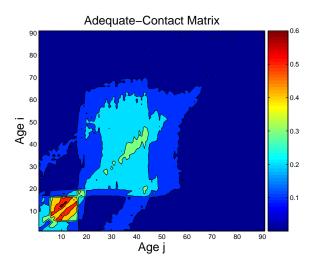


Figure 6: Adequate-contact matrix  $\beta_{ij}$ . We observe that children and teenagers have more adequate contacts than the rest of the population.

The adequate-contact matrix  $\beta_{ij}$  (figure 6) is consistent with the two blocks of mixing found in figure 3. Figure 6 shows that children and teenagers (< 20 years) are more likely to have adequate-contacts with people of their own age than the rest of the population. The average number of adequate-contacts for adults is more uniformly distributed across middle aged groups. We also observe that there is a weak coupling between middle aged adults and children.

The adequate-contact matrix  $\beta_{ij}$  can be used to obtain the forces of infection needed in a mathematical model with age structure. In order to estimate the force of infection for an specific disease, one would need to estimate the susceptibilities  $(\alpha_i)$  for each age group, the

infectivities  $(\xi_{jk})$  and the transmissibility parameter  $(\sigma)$  for the disease.

### 5 CONCLUSIONS

Contact patterns play an important role in determining the progression of epidemics. We have introduced a method for obtaining useful information on the mixing patterns of a virtual population which might lead to the spread of airborne infections. To quantify these patterns we used the social contact network generated by the epidemic simulation tool, EpiSims. We argue that mathematical models that use contact matrices based on social networks will be better able to capture age-specific infection patterns of infectious diseases than models that use transmission parameters based on homogeneous mixing or ad-hoc assumptions.

Estimating forces of infection is crucial when using models for specific infectious diseases. The forces of infection determine the rate of disease transmission and are based on the agerelated differences in the degree of mixing and contacts within the population. We used the average number of contacts and a probability distribution based on the average duration per contact to estimate an adequate-contact matrix. With the appropriate specification of disease-related parameters of susceptibility and infectivity, this matrix can be used to estimate age-dependent forces of infection for any disease.

Our results show that in general there are two main blocks of mixing within the population: children (< 20 years) and adults (> 20 years). Furthermore, we observe a weak coupling between children and middle aged adults. The average number of contacts varies among age groups, however, in general, middle aged adults have more contacts than children and older

adults.

The adequate-contact matrix in Figure 6 shows that school children are more likely to become infected than the rest of the population. This may be due to long duration of contacts children have with other children at school. In contrast, adults interact with a wider range of age groups, but their duration of contact is shorter.

One limitation of our study is that currently EpiSims does not stratify schools by classrooms and therefore the probability of having a contact with anyone in each school is the same for all children. However, children attending elementary schools mix more in their classrooms with other school children of their own age than with children of other ages. However, the EpiSims simulation does not take this factor into consideration and thus children are assumed to mix randomly with other children in their school. The spread of many childhood diseases is governed by the pattern of contact among children and therefore it is important to incorporate realistic mixing patterns. While recognizing some of the limitations in the current EpiSims simulation model, EpiSims represents a potentially powerful resource in the face of an actual outbreak.

For mathematical models of infectious diseases to be useful in guiding public health policy, they must consider age-dependent forces of infection. Individual behavior is crucial for the spread of infectious diseases and predicting disease spread is difficult. Therefore, new techniques such as the one developed here using social networks are needed as alternative tools when aggregate behavior cannot be applied to the population. The adequate-contact matrix developed here is useful in providing estimates of the age-dependent forces of infection for

mathematical models. However, much more needs to be known about the interactions between people that lead to infection before it will be possible to accurately predict an epidemic.

This research has been supported through the Mathematical Modeling and Analysis Group and the Simulation Science Group at Los Alamos National Laboratory under the Department of Energy contract W-7405-ENG-36. LAUR 05-6544.

### References

- [1] Anderson, R. M. & May, R. M. 1982 Population Dynamics of Infectious Diseases: Theory and Applications. Chapman and Hall, London.
- [2] Anderson, R. M., Gupta, S. & Ng, W. 1990 The Significance of Sexual Partner Contact Networks for the Transmission Dynamics of HIV J. AIDS 3, 417-429.
- [3] Anderson, R. M. & May, R. M. 1991 Infectious Diseases of Humans. Oxford University Press.
- [4] Barrett, C. L., Eubank, S. G. & Smith, J. P. 2005 If Smallpox Strikes Portland. Scientific American 292, 54–61.
- [5] Blythe, S. P. & Castillo-Chavez, C. 1989 Like-with-Like Preference and Sexual Mixing Models. Int. Stat. Rev. 96, 221–238.

- [6] Chowell, G., Hyman, J. M., Eubank, S. & Castillo-Chavez, C. 2003 Laws for the Movement of People Between Locations in a Large City. Phys. Rev. E 68, 1–7.
- [7] Edmunds, W. J., O'Calaghan, C. J. & Nokes, D. J. 1997 Who Mixes with Whom? A Method to Determine the Contact Patterns of Adults that may Lead to the Spread of Airborne Infections. Proc. R. Soc. Lond. B. 264, 949–957.
- [8] Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., Wang, N., & The EpiSims Team. 2004 Modeling Disease Outbreaks in Realistic Urban Social Networks. *Nature* 429, 180–184.
- [9] Grenfell, B. T. & Anderson, R. M. 1985 The Estimation of Age Related Rates of Infection from Case Notifications and Serological Data. J. Hyg. 95, 419–436.
- [10] Hethcote, H. W. & Van Ark, J. W. 1987 Epidemiological Models for Heterogeneous Populations: Proportionate Mixing, Parameter Estimation, and Immunization Programs. *Math. Biosci.* 84, 85–118.
- [11] Hethcote, H. W. & Yorke, J. A. 1984 Gonorrhea Transmission Dynamics and Control, volume 56. Springer-Verlag, New York.
- [12] Hyman, J. M. & Li, J. 1997 Disease Transmission Models with Biased Partnership Selection. Appl. Numer. Math. 24, 379–392.
- [13] Hyman, J. M., Li, J. & Stanley, E. A. 1999 The Differential Infectivity and Staged Progression Models for the Transmission of HIV. Math. Biosci. 155, 77–109.

- [14] Hyman, J. M. & Stanley, E. A. 1989 The Effect of Social Mixing Patterns on the Spread of AIDS. In Lecture notes in Biomathematics, Mathematical Approaches to Problems in Resource Management and Epidemiology, pp. 190–219. Springer-Berlin.
- [15] Jacquez, J. A., Simon, C. P., Koopman, J. S., Sattenspiel, L. & Perry, T. 1988 Modeling and Analyzing HIV Transmission: The Effect of Contact Patterns. *Math. Biosci.* 92, 119–199.
- [16] Knolle, H. 2004 A Discrete Branching Process Model for the Spread of HIV via Steady Sexual Partnerships. Math. Biol. 48, 423–443.
- [17] Koopman, J. S., Jacquez, J. A. & Park, T. S. 1989 Selective Contact within Structured Mixing with an Application to HIV Transmission Risk from Oral and Anal Sex. In Lecture notes in Biomathematics, Mathematical and Statistical Approaches to AIDS Epidemiology, pp. 317–348. Springer-Verlang, New York.
- [18] Newman, M. E. J. 2002 Spread of Epidemic Disease on Networks. Phys. Rev. E 66, 1–11.
- [19] Nold, A. 1980 Heterogeneity in Disease-Transmission Modeling. Math. Biosci. 52, 227–240.
- [20] van den Driessche, P. & Watmough, J. 2002 Reproduction Numbers and Sub-Threshold Endemic Equilibria for Compartmental Models of Disease Transmission. *Math. Biosci.* 180, 29–48.
- [21] Wallinga, J., Edmunds, W. J. & Kretzschmar, M. 1999 Perspective: Human Contact Patterns and the Spread of Airborne Infectious Diseases. *Trends Microbiol.* 7, 372–377.

[22] Zaric, G. S. 2002 Random vs. Nonrandom Mixing in Network Epidemic Models. *Health Care Management Sci.* **5**, 147–155.